Fast estimation of spatially dependent temporal vegetation trends using Gaussian Markov random fields

Fast Estimation of Spatially Dependent Temporal Vegetation Trends using Gaussian Markov Random Fields

David Bolin [a,*] Johan Lindström [a] Lars Eklundh [b] Finn Lindgren [a]

[a] *Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Box 118, SE-22100 Lund, Sweden*

[b] *Department of Physical Geography and Ecosystems Analysis, GeoBiosphere Science Centre, Lund University, Lund, Sweden*

Abstract

There is a need for efficient methods for estimating trends in spatio-temporal Earth Observation data. A suitable model for such data is a space-varying regression model, where the regression coefficients for the spatial locations are dependent. The spatial covariance structure is specified using a second order intrinsic Gaussian Markov Random Field prior. Model parameters are estimated using the Expectation Maximisation (EM) algorithm, which allows for feasible computation times for relatively large data sets. Results are illustrated with simulated data sets and real vegetation data from the Sahel area in northern Africa. The results indicate a substantial gain in accuracy compared to methods based on independent ordinary least squares regressions for the individual pixels in the data set. Use of the EM algorithm also gives a substantial performance gain over Markov Chain Monte Carlo-based estimation approaches.

Key words: Gaussian Markov Random Fields, temporal trends, vegetation data, EM algorithm, Spatio-temporal modeling, African Sahel

1 Introduction

Current awareness of global warming has directed scientific interest towards effective monitoring of land surface changes. Trends in vegetation cover are

* Corresponding author. Tel.: +46 46 2227974; fax: +46 46 2224623
Email address: bolin@maths.lth.se (David Bolin).

related to changes in climatic drivers, feedback mechanisms between the atmosphere and land surface, and human interaction. A region with rapid recent changes is the African Sahel. This zone has received much attention regarding desertification and climatic variations (Olsson, 1993; Nicholson, 2000; Lamb, 1982). The region saw an increase in rainfall in the 1960's, but since then, rainfall over the Sahel has decreased severely (Hulme, 2001). Starting in the late 1960's, the area suffered droughts for over twenty years, culminating with a severe drought and famine in 1983-1984. Since the 1983-1984 famine, and because of the growing archive of meteorological satellite observations, several authors have used satellite imagery to study vegetation in the Sahel (Tucker et al., 1985; Justice and Hiernaux, 1986; Seaquist et al., 2003).

Recently, Eklundh and Olsson (2003) observed a strong increase in seasonal vegetation index over parts of the Sahel using Advanced Very High Resolution Radiometer (AVHRR) data from the NOAA/NASA Pathfinder AVHRR Land (PAL) database (Agbu and James, 1994; James and Kalluri, 1994) for the period 1982-1999. The study was based on ordinary least squares (OLS) linear regression on individual time series extracted for each pixel in the satellite images. The changes were interpreted as a vegetation recovery from the earlier drought periods (Olsson et al., 2005), and was found to be related to an increase in rainfall (Hickler et al., 2005).

However, the time period of space observations is relatively short and the data affected by numerous disturbances, e.g. inter-sensor calibration and satellite drift (Lindström et al., 2006), atmospheric water vapour (Tanré et al., 1992), aerosol variations (Hanan et al., 1995), geometric errors, clouds, and effects of anisotropic surface reflectance (Holben, 1986; Prince and Goward, 1996). The noisy data affect significance levels of the derived trends, reducing the possibility of generating unambiguous trend images. A drawback with the OLS method is that spatial dependencies in the vegetation cover are neglected. The model presented in this work aims at generating a more efficient analysis by including these dependencies. For comparison, it is tested on the same data as were used in Eklundh and Olsson (2003).

There are several ways of incorporating the spatial context into the analysis. A simple approach is to use smoothing, either on the calculated regression coefficients, or directly on the spatio-temporal data using three-dimensional smoothing (Fan and Gijbels, 1996; Di Giacinto et al., 2005). Another approach is to construct a Bayesian Hierarchical model (Wikle and Cressie, 1998; Gelman et al., 2004) where the spatial dependence is introduced either by letting the regression errors be spatially dependent (Cressie, 1991, Chapters 2-5) or by utilising a spatially dependent prior on the regression coefficients (Besag et al., 1991; Gamerman et al., 2003).

Here, we formulate the regression based on a Bayesian hierarchical model.

The data is considered to be noisy observations of underlying fields. A prior on the regression coefficients is then introduced by assuming a spatial dependence structure for each of the underlying fields and further assuming that the time evolutions of the pixels in the fields are restricted to lie in the space spanned by the regression functions. This restriction introduces a spatial prior on the regression coefficients. The parameters of the model can, potentially, be estimated using an Markov Chain Monte Carlo (MCMC) based approach, but the large data-set makes this computationally infeasible (Bolin, 2007). Instead we estimate the model parameters using the Expectation Maximisation (EM) algorithm (Dempster et al., 1977), which allows for reasonably fast computations even when the studied data-set is large.

2 Statistical Model

Assume that we have observations in an area of size $n_1 \times n_2$ pixels at the times $t = 0, 1, \ldots, T-1$. We will from now on assume that all images are column vectorised, so that an image is represented by a vector of size $n = n_1 n_2$. Denote an observation at time $t \in [0, T-1]$ by $Y_t \in \mathbb{R}^{n-n_t}$, where n_t is the number of missing data points at time t. We assume that each observation is generated as,

$$\mathbf{Y}_t | \mathbf{X}_t, \mathbf{\Sigma}_{\epsilon_t} \in \mathsf{N}(\mathbf{A}_t \mathbf{X}_t, \mathbf{\Sigma}_{\epsilon_t}),$$

where $\mathbf{X}_t$ is an underlying field with prior distribution $\pi(\mathbf{X}_t)$, $\mathbf{\Sigma}_{\epsilon_t}$ is a measurement noise covariance matrix, and $\mathbf{A}_t$ is an observation matrix determining the observed pixels. That is, if the field is measured at the locations $i_1, \ldots, i_{n-n_t}$, the observation matrix, $\mathbf{A}_t$, will be of size $(n - n_t) \times n$, and have all elements equal to zero except for

$$A_{1,i_1} = \ldots = A_{n-n_t, i_{n-n_t}} = 1.$$

Assuming that $\mathbf{X}_{t_1}$ and $\mathbf{X}_{t_2}$ are independent for $t_1 \neq t_2$, the probability density for $\mathbf{X}^\top = (\mathbf{X}_0^\top, \ldots, \mathbf{X}_{T-1}^\top)^\top$ would be

$$\pi(\mathbf{X}) = \prod_{t=0}^{T-1} \pi(\mathbf{X}_t). \tag{1}$$

To estimate time varying trends in the observations, we introduce a $T \times m$-matrix $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_m]$ containing basis functions for the trends. These basis functions are chosen to be linearly independent, i.e. $\text{rank}(\mathbf{F}) = m$. Restricting $\mathbf{X}$ to follow these trends, we get

$$\mathbf{X} = (\mathbf{F} \otimes \mathbf{I})\mathbf{K}, \tag{2}$$

where $\mathbf{K} = [\mathbf{K}_1^\top, \ldots, \mathbf{K}_m^\top]^\top$ contains the coefficients for the trends and $\mathbf{I}$ is an $n \times n$ identity-matrix. The prior distribution for $\mathbf{K}$ is obtained by evaluating

(1) conditionally on the restriction (2).

2.1 Gaussian Markov Random Fields

A suitable prior describing the spatial dependencies in $\mathbf{X}_t$ is a Gaussian Markov Random Field (GMRF) (see Rue and Held, 2005, for extensive details). A random variable $\mathbf{x} = (x_1, \ldots, x_n)^\top \in \mathsf{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1})$ is called a GMRF if the joint distribution for $\mathbf{x}$ satisfies $\pi(x_i|x_{-i}) = \pi(x_i|x_{\mathcal{N}_i})\ \forall\ i$. Here $\mathcal{N}_i$ is the neighbourhood of i and x_{-i} denotes all elements in $\mathbf{x}$ except x_i. An important implication of this is that if $i \neq j$ then:

$$x_i \perp x_j | x_{-\{i,j\}} \iff Q_{i,j} = 0 \iff j \notin \mathcal{N}_i. \tag{3}$$

This mean that the following properties are equivalent: 1) x_i and x_j are conditionally independent 2) the corresponding element in the precision matrix, $Q_{i,j}$ is zero and 3) i and j are not neighbours. Since $Q_{i,j} \neq 0$ only if i and j are neighbours, most GMRFs will have sparse precision matrices. The sparse precision matrix is the main reason for using GMRFs in this work; none of the following computations would be feasible for full matrices because of the large number of observations and nodes in the fields.

In the analysis, we will use a special type of GMRFs called Intrinsic Gaussian Markov Random Fields (IGMRFs). An IGMRF is improper, that is, its precision matrix does not have full rank.

Definition 1 *Let* $\mathbf{x} = (x_1, \ldots, x_n)^\top$*, and let* $\mathbf{Q}$ *be an* $n \times n$ *symmetric positive semi-definite matrix with rank* $n - k$ *such that*

$$Q_{i,j} \neq 0 \iff j \in \mathcal{N}_i.$$

Then $\mathbf{x}$ *is an Intrinsic GMRF, with parameters* $\boldsymbol{\mu}$ *and* $\mathbf{Q}$ *if its density is*

$$\pi(\mathbf{x}) = \frac{|\mathbf{Q}|_*^{\frac{1}{2}}}{(2\pi)^{\frac{n-k}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{Q}(\mathbf{x} - \boldsymbol{\mu})\right). \tag{4}$$

Here, $|\mathbf{Q}|_*$ denotes the product of the $n - k$ non-zero eigenvalues of $\mathbf{Q}$, and following Rue and Held (2005), we will call this the generalised determinant of $\mathbf{Q}$. It should be noted that the density of an IGMRF is not a true density since it is invariant to additions of vectors from the null-space of $\mathbf{Q}$. Hence, it cannot be normalised to integrate to one. The parameters, $\boldsymbol{\mu}$ and $\mathbf{Q}$, no longer represent the mean and precision, since these formally do not exist; however, we will for convenience continue denoting them as mean and precision.

We choose a second-order polynomial IGMRF (Gamerman et al., 2003; Rue and Held, 2005, Section 3.4.2) as a smooth prior. The field is invariant to the addition of an arbitrary plane, $p(i,j) = a + bi + cj$, and the precision matrix, $\kappa\mathbf{Q}$ has a rank deficiency of 3. Here κ is a scaling parameter that governs the strength of the dependence in the IGMRF. The second order IGMRF can be seen as a set of penalties on

$$(\Delta^2_{(1,0)} + \Delta^2_{(0,1)})x_{i,j}, \tag{5}$$

where $\Delta_{(1,0)}$ and $\Delta_{(0,1)}$ are forward-differences in the directions $(1,0)$ and $(0,1)$ respectively, and $\Delta^2_{(1,0)}+\Delta^2_{(0,1)}$ is a discrete approximation of the Laplace operator. Due to (5) this IGMRF can be interpreted as penalties on the second derivatives if the field is used to model an underlying continuous field.

2.2 Regression Model

Using the second order IGMRF, described above, a distribution for each $\mathbf{X}_t$ in (1), with an unknown precision parameter κ, we get that $\mathbf{X}|\kappa \in \mathsf{N}(0, (\kappa\overline{\mathbf{Q}})^{-1})$, where $\overline{\mathbf{Q}} = \mathbf{I} \otimes \mathbf{Q}_X$ and $\mathbf{Q}_X$ is the precision matrix for a second order IGMRF. Given the restriction in (2), we can now derive the distribution of the parameter field, $\mathbf{K}$.

Proposition 2 *$\mathbf{K}|\kappa$ is an IGMRF of rank $m(n-3)$ with zero mean and covariance matrix $(\kappa\mathbf{Q})^{-1}$, where $\mathbf{Q} = (\mathbf{F}^\top\mathbf{F}) \otimes \mathbf{Q}_X$.*

For proof see Appendix A.

Note that the sparsity structure of $\mathbf{Q}$ is partially determined by the orthogonality of the regression basis, $\mathbf{F}$, see Figure 1.

Using the distribution for $\mathbf{K}|\kappa$ from Proposition 2, we can now write the observations, $\mathbf{Y}$, as $\mathbf{Y}|\mathbf{K}, \mathbf{\Sigma}_\epsilon \in \mathsf{N}(\mathbf{AK}, \mathbf{\Sigma}_\epsilon)$. Here, $\mathbf{\Sigma}_\epsilon$ is a block diagonal matrix such that

$$\mathbf{\Sigma}_\epsilon = \begin{bmatrix} \mathbf{\Sigma}_{\epsilon_0} & & 0 \\ & \ddots & \\ 0 & & \mathbf{\Sigma}_{\epsilon_{T-1}} \end{bmatrix},$$

and $\mathbf{A} = \mathrm{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_{T-1})(\mathbf{F} \otimes \mathbf{I})$, where $\mathbf{I}$ is an $n \times n$ identity matrix. The posterior distribution for $\mathbf{K}$ given data, $\mathbf{\Sigma}_\epsilon$, and κ becomes

$$\begin{aligned} \pi(\mathbf{K}|\mathbf{Y}, \mathbf{\Sigma}_\epsilon, \kappa) &\propto \pi(\mathbf{Y}|\mathbf{K}, \mathbf{\Sigma}_\epsilon)\pi(\mathbf{K}|\kappa)\pi(\kappa)\pi(\mathbf{\Sigma}_\epsilon) \\ &\propto \pi(\mathbf{Y}|\mathbf{K}, \mathbf{\Sigma}_\epsilon)\pi(\mathbf{K}|\kappa). \end{aligned}$$

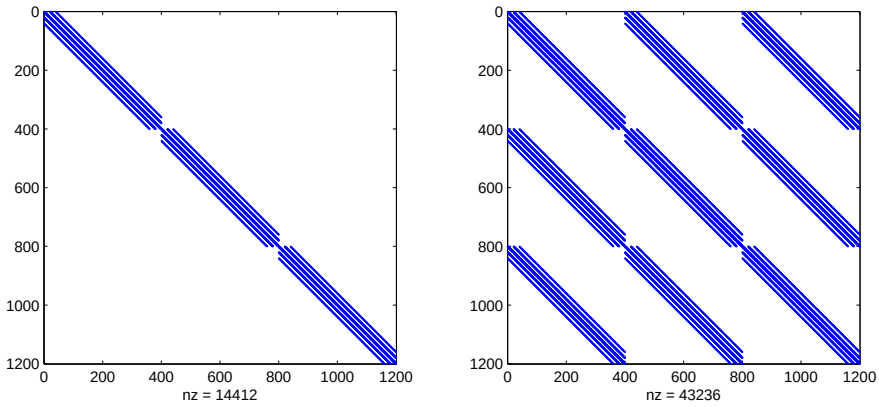

Fig. 1. The figure depicts the sparsity structure in $\mathbf{Q}$, Proposition 2, for a field with 400 nodes and three regression basis vectors for orthogonal regression basis (left) and non-orthogonal regression basis (right) respectively. The number of non-zero elements in each of the matrices is denoted nz.

This follows since we choose to use flat priors for κ and $\mathbf{\Sigma}_\epsilon$, i.e. $\pi(\kappa) \propto 1$ and $\pi(\mathbf{\Sigma}_\epsilon) \propto 1$. It can be shown (Rue and Held, 2005, p. 39) that the posterior distribution of $\mathbf{K}$ is $\mathbf{K}|\mathbf{Y}, \mathbf{\Sigma}_\epsilon, \kappa \in \mathsf{N}(\boldsymbol{\mu}_{K|Y}, \mathbf{Q}_{K|Y}^{-1})$, with

$$\boldsymbol{\mu}_{K|Y} = \mathbf{Q}_{K|Y}^{-1}\mathbf{A}^\top\mathbf{\Sigma}_\epsilon^{-1}\mathbf{Y} \qquad \text{and} \qquad \mathbf{Q}_{K|Y} = \kappa\mathbf{Q} + \mathbf{A}^\top\mathbf{\Sigma}_\epsilon^{-1}\mathbf{A}. \tag{6}$$

2.3 Noise Model

To complete the model we need a noise model, i.e. we need to determine the structure of $\mathbf{\Sigma}_\epsilon$. Many of the factors that the measurement noise should model are local phenomena, such as aerosol and cloud cover. Since it seems unreasonable that the scale of these disturbances would be the same over the entire, large, region, we assume a noise model with a different noise variance at each pixel. Further, for simplicity, we take the noise to be spatially uncorrelated. Thus $\mathbf{\Sigma}_\epsilon$ is a $(nT - \sum_t n_t) \times (nT - \sum_t n_t)$ diagonal matrix with n unique diagonal elements, $\sigma_1^2, \ldots, \sigma_n^2$, which represent the noise variances of the n different pixels. Note that $\mathbf{\Sigma}_\epsilon$ can be constructed by starting from the matrix $\mathbf{I} \otimes \text{diag}(\sigma_1^2 \ \ldots \ \sigma_n^2)$, where $\mathbf{I}$ is a $T \times T$ identity-matrix, and then removing all rows and columns that correspond to missing observations.

3 EM Parameter Estimation

By construction, the structure of the precision matrix, $\mathbf{Q}$, in (6) is known a priori, whereas the parameters κ and $\mathbf{\Sigma}_\epsilon$ have to be estimated. These parameters can, potentially, be estimated using an MCMC based approach. However,

two difficulties arise when attempting to estimate this model with MCMC. Firstly, given the large number of variance parameters in $\mathbf{\Sigma}_\epsilon$, parameters that are likely to be correlated, we find it very challenging to construct an efficient proposal distribution that ensures good mixing of the Markov chain. Secondly, even if this issue is avoided by assuming a simpler measurement noise model with one common variance parameter for all the pixels in the field, the large data-set makes an MCMC based estimation approach computationally infeasible (see Bolin, 2007, for details).

A better alternative is to interprete the problem as a missing data problem. This interpretation facilitates use of the EM algorithm, allowing for much faster computations of maximum likelihood parameter estimates (a comparission of computational times for simulated data is given in the last paragraph of Section 6.1).

Augmenting the observed data, $\mathbf{Y}$, with the unknown regression coefficients, $\mathbf{K}$, we obtain the complete data, $(\mathbf{Y}, \mathbf{K})$, and the augmented likelihood becomes

$$L(\kappa, \mathbf{\Sigma}_\epsilon | \mathbf{Y}, \mathbf{K}) = \pi(\mathbf{Y}, \mathbf{K} | \kappa, \mathbf{\Sigma}_\epsilon) = \pi(\mathbf{Y} | \mathbf{K}, \mathbf{\Sigma}_\epsilon)\pi(\mathbf{K}|\kappa).$$

Taking the parameters as, $\theta = (\kappa, \mathbf{\Sigma}_\epsilon)$, the loss-function is

$$\mathcal{Q}(\theta, \theta^{(i)}) = \mathsf{E}(\log(L(\theta|\mathbf{K}, \mathbf{Y}))|\mathbf{Y}, \theta^{(i)}), \tag{7}$$

where $\theta^{(i)}$ is an estimate of θ at iteration i, and the expectation is taken over $\mathbf{K}$.

The likelihood consists of two parts, $\pi(\mathbf{Y}|\mathbf{K}, \mathbf{\Sigma}_\epsilon)$ and $\pi(\mathbf{K}|\kappa)$. For the first part we have that $(\mathbf{Y}|\mathbf{K}, \mathbf{\Sigma}_\epsilon) \in \mathsf{N}(\mathbf{AK}, \mathbf{\Sigma}_\epsilon)$, hence

$$\begin{aligned}
\log(\pi(\mathbf{Y}|\mathbf{K}, \mathbf{\Sigma}_\epsilon)) &= \log\left(\frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{\Sigma}_\epsilon|^{\frac{1}{2}}}\right) - \frac{1}{2}(\mathbf{Y} - \mathbf{AK})^\top \mathbf{\Sigma}_\epsilon^{-1}(\mathbf{Y} - \mathbf{AK}) \\
&= -\frac{1}{2}\sum_{j=1}^{n}\left(n_j \log(\sigma_j^2) + \frac{1}{\sigma_j^2}\sum_{k=1}^{n_j}(\mathbf{Y} - \mathbf{AK})_{j_k}^2\right) + \text{const.}
\end{aligned}$$

Here, n_j is the number of observations at pixel j (not to be confused with n_t, the number of missing observations at time t), and the last sum runs over these observations, with $(\mathbf{Y} - \mathbf{AK})_{j_k}$ referring to the deviation of the k^{th} observation of pixel j, from the regression line. Similarly, since $\mathbf{K}|\kappa \in \mathsf{N}(0, (\kappa\mathbf{Q})^{-1})$, we get

$$\log(\pi(\mathbf{K}|\kappa)) = \frac{1}{2}\log(|\kappa\mathbf{Q}|_*) - \frac{\kappa}{2}\mathbf{K}^\top\mathbf{QK} + \text{const.} \tag{8}$$

By Proposition 2, the rank of $\mathbf{Q}$ is $m(n-3)$, where n is the number of pixels in the image, giving

$$\log(\pi(\mathbf{K}|\kappa)) = \frac{m(n-3)}{2}\log(\kappa) - \frac{\kappa}{2}\mathbf{K}^\top\mathbf{QK} + \text{const.} \tag{9}$$

Thus, the loss-function becomes

$$\begin{aligned}\mathcal{Q}(\theta,\theta^{(i)}) = \frac{m(n-3)}{2}\log(\kappa) - \frac{\kappa}{2}\mathsf{E}(\mathbf{K}^\top\mathbf{Q}\mathbf{K}|*) \\ - \frac{1}{2}\sum_{j=1}^{n}\left(n_j\log(\sigma_j^2) + \frac{1}{\sigma_j^2}\sum_{k=1}^{n_j}\mathsf{E}((\mathbf{Y}-\mathbf{A}\mathbf{K})_{j_k}^2|*)\right) + \text{const.},\end{aligned} \tag{10}$$

where the notation $\mathsf{E}(\dots|*)$ means $\mathsf{E}(\dots|\mathbf{Y},\kappa^{(i)},\mathbf{\Sigma}_\epsilon^{(i)})$.

To calculate $\text{argmax}_\theta\,\mathcal{Q}(\theta,\theta^{(i)})$, we differentiate (10) with respect to the parameters κ and $\sigma_1^2,\dots,\sigma_n^2$ and set these derivatives equal to zero, yielding

$$\begin{aligned}\sigma_j^{2(i+1)} &= \frac{1}{n_j}\sum_{k=1}^{n_j}\mathsf{E}((\mathbf{Y}-\mathbf{A}\mathbf{K})_{j_k}^2|*),\ \text{for } 1\le j\le n, \text{ and} \\ \kappa^{(i+1)} &= \frac{m(n-3)}{\mathsf{E}(\mathbf{K}^\top\mathbf{Q}\mathbf{K}|*)}.\end{aligned} \tag{11}$$

Recall that these updating equations are derived under the assumption of flat priors for κ and $\sigma_1^2,\cdots,\sigma_n^2$. However it can easily be shown that Gamma and independent inverse-Gamma priors for κ and $\sigma_1^2,\cdots,\sigma_n^2$ respectively will yield tractable updating expressions in (11).

In general the expectations in (11) can now be found by simulating from the posterior density, $\pi(\mathbf{K}|\mathbf{Y},\theta^{(i)})$, using a Monte Carlo-approach. However, in this case the observations given the underlying field, $\pi(\mathbf{Y}|\mathbf{K},\mathbf{\Sigma}_\epsilon^{(i)})$, are Gaussian which in turn implies that the posterior density, $\pi(\mathbf{K}|\mathbf{Y},\theta^{(i)})$, is Gaussian and we can calculate the updating rules in (11) analytically. The first expectation in (11) becomes

$$\begin{aligned}\mathsf{E}((\mathbf{Y}-\mathbf{A}\mathbf{K})_{j_k}^2|*) &= Y_{j_k}^2 - 2Y_{j_k}\mathbf{A}_{(j_k,\bullet)}\mathsf{E}(\mathbf{K}|*) + \mathsf{E}((\mathbf{A}_{(j_k,\bullet)}\mathbf{K})^2|*) \\ &= (Y_{j_k} - \mathbf{A}_{(j_k,\bullet)}\mathsf{E}(\mathbf{K}|*))^2 + \mathbf{A}_{(j_k,\bullet)}\mathsf{V}_{K|*}\mathbf{A}_{(j_k,\bullet)}^\top.\end{aligned} \tag{12}$$

Here, $\mathbf{A}_{(j_k,\bullet)}$ denotes the row in $\mathbf{A}$ corresponding to the k^{th} observation of pixel j, and $\mathsf{E}(\mathbf{K}|*) = \boldsymbol{\mu}_{K|Y}$ and $\mathsf{V}_{K|*} = \mathbf{Q}_{K|Y}^{-1}$ are given by (6).

Using that both the expected value and the trace of a matrix are linear operators, and the cyclic property of the trace, $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CAB})$, we can calculate $\mathsf{E}(\mathbf{K}^\top\mathbf{Q}\mathbf{K}|*)$ as

$$\begin{aligned}\mathsf{E}(\mathbf{K}^\top\mathbf{Q}\mathbf{K}|*) &= \text{tr}(\mathsf{E}(\mathbf{K}^\top\mathbf{Q}\mathbf{K}|*)) = \text{tr}(\mathbf{Q}\mathsf{E}(\mathbf{K}\mathbf{K}^\top|*)) \\ &= \text{tr}(\mathbf{Q}(\mathbf{Q}_{K|Y}^{-1} + \mathsf{E}(\mathbf{K}|*)\mathsf{E}(\mathbf{K}|*)^\top)) \\ &= \text{tr}(\mathbf{Q}\mathbf{Q}_{K|Y}^{-1}) + \mathsf{E}(\mathbf{K}|*)^\top\mathbf{Q}\mathsf{E}(\mathbf{K}|*).\end{aligned} \tag{13}$$

Calculating $\text{tr}(\mathbf{Q}\mathbf{Q}_{K|Y}^{-1})$ and $\mathbf{A}_{(j_k,\bullet)}\mathbf{Q}_{K|Y}^{-1}\mathbf{A}_{(j_k,\bullet)}^\top$ in (12) and (13) might not seem feasible for large fields. However, three things should be noted: Firstly

$\mathbf{A}_{(j_k,\bullet)}\mathbf{Q}_{K|Y}^{-1}\mathbf{A}_{(j_k,\bullet)}^{\top} = \mathrm{tr}(\mathbf{A}_{(j_k,\bullet)}^{\top}\mathbf{A}_{(j_k,\bullet)}\mathbf{Q}_{K|Y}^{-1})$. Secondly due to the trace operator only the diagonal elements of the products $(\bullet\mathbf{Q}_{K|Y}^{-1})$ need to be calculated. Thirdly $\mathbf{Q}$ and $\mathbf{A}_{(j_k,\bullet)}^{\top}\mathbf{A}_{(j_k,\bullet)}$ are at least as sparse as $\mathbf{Q}_{K|Y}$, thus to calculate the traces we will *at most* need the elements of $\mathbf{Q}_{K|Y}^{-1}$ that correspond to neighbouring points in the GMRF, given the Cholesky factor these elements can be calculated *without calculating the entire inverse* (Rue and Martino, 2007), making (12) and (13) computationally feasible.

4 Testing for Significant Trends

A relevant question to ask is where changes in the vegetation have occurred over the course of the studied time period. A method which has been used to answer this question for the Sahel data set is to find the significant trends in the data (Eklundh and Olsson, 2003). For an OLS regression this is relatively straightforward, although one has to keep in mind that the obtained significant trends describe whether trends at each pixel are individually significant, and not if the trends for all pixels in an entire region are significant. To obtain the latter one has to consider techniques for multiple hypothesis testing, one option would be to do hypothesis testing on the contour lines of the K_i estimates. How to do such tests is, however, far from trivial, and will not be investigated here.

Using the GMRF model and parameter estimation algorithm described above, maximum-likelihood estimates of κ and $\sigma_1^2, \cdots, \sigma_n^2$ are obtained. Given these the conditional posterior for the $\mathbf{K}$-field is

$$\mathbf{K}|\mathbf{Y}, \kappa, \sigma_1^2, \cdots, \sigma_n^2 \in \mathsf{N}\Big(\boldsymbol{\mu}_{K|Y}, \mathbf{Q}_{K|Y}^{-1}\Big),$$

with $\boldsymbol{\mu}_{K|Y}$ and $\mathbf{Q}_{K|Y}$ defined in (6).

The conditional variances of the $\mathbf{K}$-estimates are given by the diagonal elements in $\mathbf{Q}_{K|Y}^{-1}$. Let k_i^j be the coefficient corresponding to the trend $\mathbf{f}_j$ at pixel x_i, and let $\hat{\sigma}_i^j$ denote the corresponding standard deviation given by the square root of the relevant diagonal element in $\mathbf{Q}_{K|Y}^{-1}$. A simple hypothesis test would now be to reject the null hypothesis $H_0 : k_i^j = 0$ against $H_1 : k_i^j \neq 0$, at a 95% confidence level if

$$\left|\frac{k_i^j}{\hat{\sigma}_i^j}\right| > \lambda_{0.025},$$

where $\lambda_{0.025}$ is the 2.5%-quantile of the $\mathsf{N}(0,1)$-distribution.

Simulation studies (see Table 1) indicate that this test rejects the null hypothesis too often, i.e. the corresponding confidence interval for k_i^j is too small. The

problem is that the uncertainty in the estimated parameters κ and $\sigma_1^2, \cdots, \sigma_n^2$ are ignored. A solution to this problem is to instead, as in OLS regression, reject the null hypothesis if

$$\left| \frac{k_i^j}{\hat{\sigma}_i^j} \right| > t_{0.025}(f),$$

where $t_{0.025}(f)$ is the 2.5%-quantile of the Student's t-distribution with f degrees of freedom. To perform this test, we have to determine f, which in OLS depends on the number of data points used to estimate k_i^j, and thus $\hat{\sigma}_i^j$. Since κ determines the smoothness in the parameter field, a large κ-value would imply that the regression parameters are highly correlated. In this case, the estimate of a parameter at a specific pixel is influenced, not only by the observations at that pixel, but also by observation of surrounding pixels. On the other hand, if $\kappa = 0$, the model breaks down to n independent OLS regressions, and each parameter is estimated using only the observations of the corresponding pixel. Thus, it is reasonable to assume that f will be dependent on κ. It is, however, not easy to determine an analytic expression for f, and we therefore use a simulation based algorithm for estimating the degrees of freedom: Given estimated values of κ and $\mathbf{\Sigma}_\epsilon$, we simulate new $\mathbf{K}$-fields and corresponding data sets. From these data, we use the EM algorithm to calculate new ML-estimates, $\kappa^*, \Sigma_\epsilon^*$, and new posterior means $\mathbf{K}^* = \mathsf{E}(\mathbf{K}|\mathbf{Y}, \kappa^*, \mathbf{\Sigma}_\epsilon^*)$. The residuals, $\boldsymbol{\varepsilon} = \sum_j \mathbf{f}_j^\top \otimes (\mathbf{K}_j - \mathbf{K}_j^*)$, are calculated and each residual, $\varepsilon_{i,t}$, is normalised by division with its corresponding standard deviation $\mathsf{V}(\sum_j f_j(t) K_j^*(i))^{1/2}$. In standard OLS regression, the residuals would now follow a Student's-t distribution with unknown degrees of freedom. Thus we estimate f by maximising the log-likelihood, of the t-distribution $\sum_{j,t} \log t_f(\varepsilon_{j,t})$, with respect to f, using a simplex search method (Lagarias et al., 1998). Simulation studies, Table 1, indicate that (4) with f estimated as described above gives correct significance levels.

5 Data

Many Earth Observation studies use Normalised Difference Vegetation Index (NDVI) data (Rouse et al., 1973) as a remotely sensed measure of ground vegetation. The principle behind the NDVI is that leaf chlorophyll absorbs red light, whereas the cell structure of leaves reflects light in the near-infrared spectra. With these properties in mind, NDVI is calculated as

$$\text{NDVI} = \frac{R_{\text{NIR}} - R_{\text{RED}}}{R_{\text{NIR}} + R_{\text{RED}}}, \tag{14}$$

where R_{NIR} and R_{RED} are the measured reflectancies in the near-infrared and red spectral bands, respectively. Thus, dense vegetation tends to have positive

NDVI values, while soil and other areas with little vegetation tend to have NDVI values close to zero. Many studies have confirmed that NDVI correlates well with important vegetation variables, e.g. biomass, and vegetation greenness (Myneni et al., 1997; Sellers et al., 1997).

The Sahel is a semi-arid region, and it should be noted that NDVI data is well suited for studies of such areas; because of the low humidity, the atmospheric contamination of the data is fairly low (Chappell et al., 2001), and because of the sparse vegetation, saturated NDVI values are less of a problem than for areas with denser vegetation cover (Prince, 1991).

5.1 Data Preprocessing

The PAL database is derived from measurements by the AVHRRs on-board the NOAA series of meteorological satellites (Kidwell, 1998). Before the NDVI values are calculated, the AVHRR data undergoes a number of calibration steps (see Agbu and James, 1994, for details). The data is then mapped to 8 km $\times$ 8 km pixels and 10-day maximum value composites are generated to compensate for negative bias due to clouds (Holben, 1986). In spite of these corrections, the data still contains unwanted noise.

The end result of the above processing is a set of 10-day maximum value composites, for a total of 36 values per year. However we are interested in assessing the year-to-year change in vegetation growth. Therefore the seasonal NDVI integral which gives a measure of the total vegetation growth during each year is of interest.

In order to generate annual data, we use the Savitzky and Golay (1964) algorithm of the TIMESAT processing scheme (Jönsson and Eklundh, 2002, 2004). The method begins by estimating the number of annual growing seasons and the corresponding season lengths. When the approximate onsets and ends of the growing seasons have been found, the time series is smoothed using a robust locally weighted least-squares fit of a quadratic polynomial to the upper envelope of the NDVI data (Press et al., 1992). The fit is adapted to the envelope of the data to account for negatively biased noise, and the weights in the least-squares fitting are determined using cloud cover data from the NOAA/NASA Cloud AVHRR (CLAVR) data set (Stowe et al., 1991). To obtain annual data, the seasonal integrals are calculated from the fitted functions and used in the following analysis. To further reduce the amount of noise in the data, all seasons which have more than one missing observation are removed. Also, to avoid artificial growing seasons in desert areas, all seasons which have less than two NDVI values above 0.1 are set to zero.

6 Results

Recall that the model presented in Section 2.2 (from now on referred to as the GMRF model) is to be used for estimating trends in vegetation data. In this section, the model is compared to simple independent OLS regression for each pixel (from now on referred to as the OLS model) using simulated data and real NDVI data.

6.1 Simulated Data

The two models are compared using 17 years of simulated data. For these tests, we use two orthogonal trend basis functions, one constant and one linearly increasing. Hence, $\mathbf{K}_1$ will contain the intercepts, and $\mathbf{K}_2$ the slopes for the linear trends.

To create a spatially dependent data set, $\mathbf{Y} = (\mathbf{Y}_0, \ldots, \mathbf{Y}_{16})$, we generate dependent $\mathbf{K}_1$- and $\mathbf{K}_2$-fields from the distribution in Proposition 2, with $\kappa = 0.5$. To create a data set without spatial dependencies, we instead generate $\mathbf{K}_1$ and $\mathbf{K}_2$ by independently drawing values from $\mathsf{N}(0, 1)$. In both cases, $\mathbf{Y}_t$, $t \in [0, 16]$, is created as $\mathbf{Y}_t = f_1(t)\mathbf{K}_1 + f_2(t)\mathbf{K}_2 + \epsilon$, where ϵ is a draw from $N(0, \Sigma_{\epsilon_t})$ and $\mathbf{\Sigma}_{\epsilon_t}$ is defined as in Section 2.3, with pixel variances σ_i^2 drawn independently from a uniform distribution on $[0, \alpha]$. We use images of size 40×40 pixels in four different cases:

D1 Spatially dependent data with $\kappa = 0.5$ and $\alpha = 1$.
D10 Spatially dependent data with $\kappa = 0.5$ and $\alpha = 10$.
R1 Random data ($\kappa = 0$) with $\alpha = 1$.
R10 Random data with $\alpha = 10$.

The accuracy of the models are compared using the following tests:

(1) Calculate the percentage of pixels correctly labeled as significant trends, i.e. the number of significant trends with correct sign divided by the total number of pixels.
(2) Estimate the number of times that the confidence intervals for the estimated trend coefficients cover the actual value. This number should be close to 95%.
(3) Compare the estimated $\mathbf{K}$-fields with the actual $\mathbf{K}$-fields using the Frobenius norm: $K_i^\varepsilon = \|\mathbf{K}_i - \mathbf{K}_i^*\|_F$.
(4) Compare the $\mathbf{\Sigma}_\epsilon$ estimates with the actual measurement noise variances: $\Sigma^\varepsilon = \|\mathbf{\Sigma}_\epsilon - \mathbf{\Sigma}_\epsilon^*\|_F$.

Each of the tests are done on 100 different data sets for each of the four cases

described above. The results for the tests are summarised in Table 1. As seen in the table, the two models are fairly similar for random data whereas the GMRF-estimations of the **K**-fields are more accurate for spatially dependent data. This can be seen in Figure 2 where generated **K**-fields are shown together with the two estimates. The lower variance of the estimates also results in a higher percentage of correctly labeled significant trends for spatially dependent data. The GMRF algorithm is implemented in C/C++ using GMRFLib (Rue, 2007). For these simulations, the average time for each step in the EM algorithm was 0.26 seconds, which, for example, resulted in an average computation time of 9.80 seconds for the **D1** simulations. The computations were performed on a dual CPU (2×2.1GHz) personal computer. For comparison, an MCMC based estimation algorithm was also implemented in C/C++, assuming a simpler measurement noise model with one common variance parameter for all the pixels in the field (see Bolin, 2007, for details). The average computation time for one iteration in this algorithm was 0.11 seconds, which resulted in a computation time of 1050 seconds for parameter estimates based on 10^4 iterations.

		K_1^ε	K_2^ε	Σ^ε	Cov (%)	Sig (%)
D1	OLS	6.84	1.39	8.40	95.01	94.73
	GMRF	3.63	0.74	8.09	94.87	97.12
D10	OLS	21.66	4.42	84.43	95.06	84.40
	GMRF	6.27	1.28	79.86	94.85	95.41
R1	OLS	6.83	1.40	8.38	94.96	94.43
	GMRF	6.81	1.43	11.92	94.88	94.42
R10	OLS	21.63	4.45	83.75	94.90	82.78
	GMRF	21.02	4.57	84.57	94.44	82.58

Table 1
Results for simulated data. Each value is the average over 100 different runs. Here, K_1^ε, K_2^ε, and Σ^ε show the difference, measured in the Frobenius norm, of the true and estimated $\mathbf{K}_1$- , $\mathbf{K}_2$- and Σ_ϵ-fields, respectively. Cov is the estimated coverage percentage for the $\mathbf{K}_2$ confidence intervals, which should be close to the nominal 95%. Sig is the percentage of correctly labeled significant trends.

6.2 Sahel Data

We base the analysis on the seasonal NDVI integrals of the 18 years of data. The results from the previous section indicates that more accurate vegetation trend estimates can be obtained using the GMRF-model instead of independent OLS regression for each pixel.

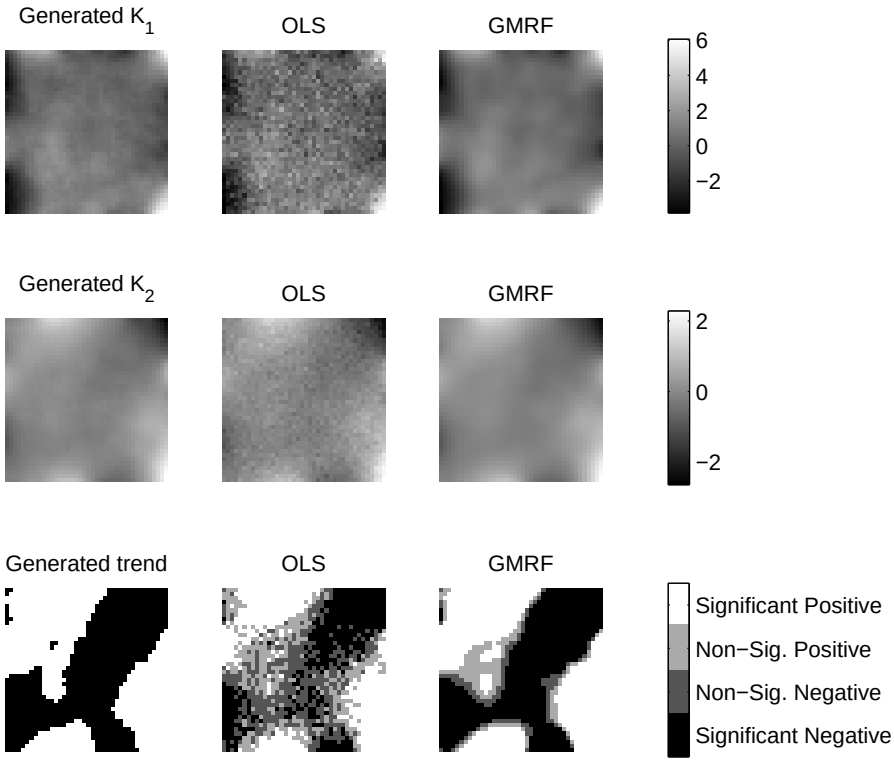

Fig. 2. Results from a simulated data set with $\kappa = 0.5$ and $\alpha = 10$. The two upper rows show the true $\mathbf{K}_1$- and $\mathbf{K}_2$-fields and the estimates generated by the OLS model and the GMRF model. Notice that the GMRF estimates are smoother and more accurate. On the bottom row, the left figure indicates where the true $\mathbf{K}_2$-field is positive (red) and negative (blue). The other two figures show the significance estimates for the $\mathbf{K}_2$-trends. Notice that more correct significant trends are found by the GMRF model.

The GMRF model assumes as common precision parameter for the entire field, which is probably not valid for the entire Sahel. Therefore, the following results were obtained by dividing the area into smaller subareas, with a (large) overlap between the subareas, and applying the GMRF model to each of these subareas. This way, κ is assumed to be constant only on the smaller subareas, and the overlap between neighbouring subareas ensure that no discontinuities appear in the joints between the subareas.

Using two orthogonal trend basis functions, one constant, and one linearly increasing, we obtain estimates shown in Figures 3 and 4. In the figures, we see that the estimates are somewhat similar, but that the GMRF estimates are much smoother than the OLS estimates. A big difference in the estimates can be seen in Figure 5. Here, the areas with significant linear trends are shown, and we see that, compared to the OLS, the GMRF estimate has more significant trends and larger contiguous regions with significant trends. This result is expected since the GMRF estimate takes advantage of the spatial

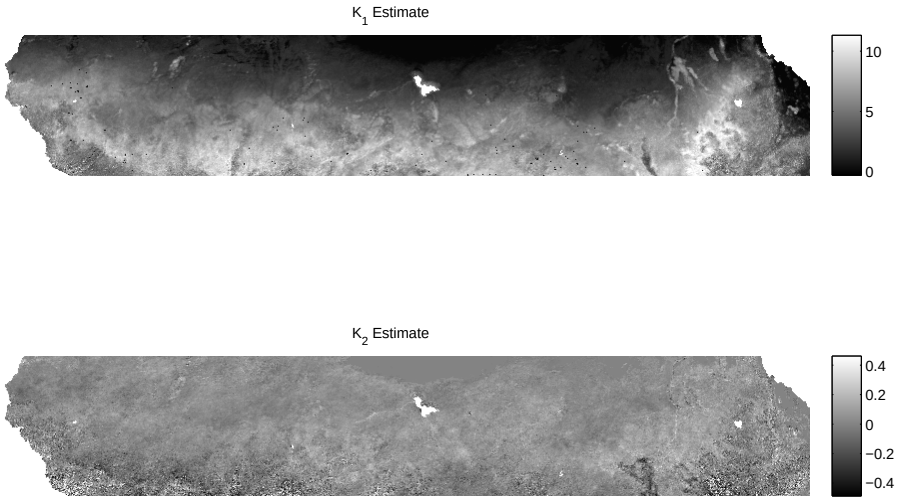

Fig. 3. OLS estimates of the intercepts (upper figure) and slopes (lower figure) for the Sahel data. Notice that especially the $\mathbf{K}_2$ estimate seems to contain a large amount of noise.

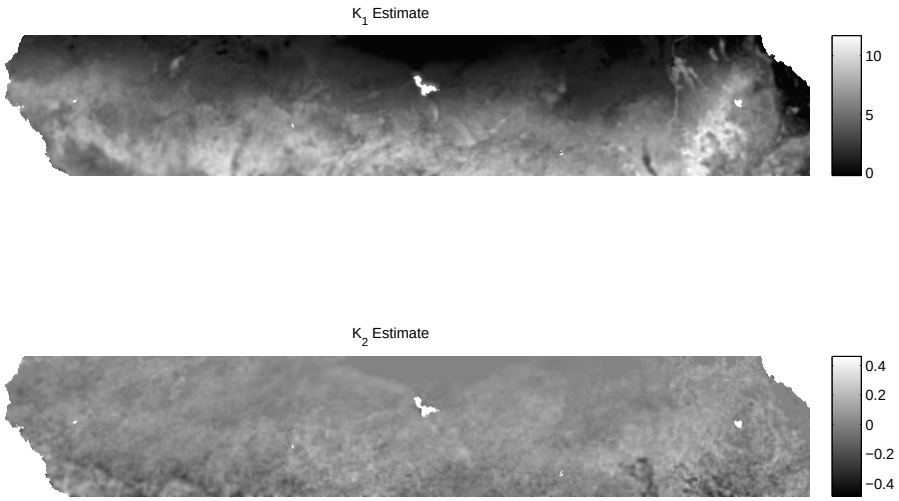

Fig. 4. GMRF estimates of the intercepts (upper figure) and slopes (lower figure) for the Sahel data. Notice that both these estimates are much smoother than the corresponding estimates in Figure 3.

dependencies in the data, and therefore uses more observations for each point estimate. As indicated in Section 4, contour plots of the $\mathbf{K}_i$ estimates might also be of interest for visualisation and hypothesis testing. In Figure 6, a

contour plot of the GMRF K_2 estimate is shown. In this figure, it is easy to see where large increases in vegetation has occurred. This is a big advantage with the GMRF method, since similar contour plots for the OLS estimate are too noisy to easily interpret. The larger contiguous regions and smoother estimates will most likely aid future interpretation of the data and make it easier to detect underlying reasons for the detected changes in vegetation.

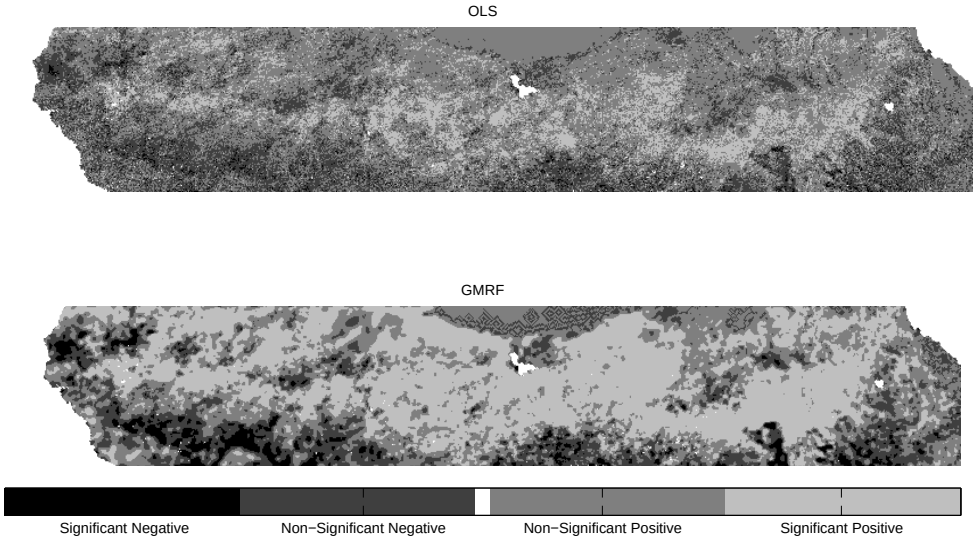

Fig. 5. Significance estimates for the slope of the linear trends using the OLS model (upper figure) and the GMRF model (lower figure). Notice the large number of significant positive linear trends in the GMRF estimate.

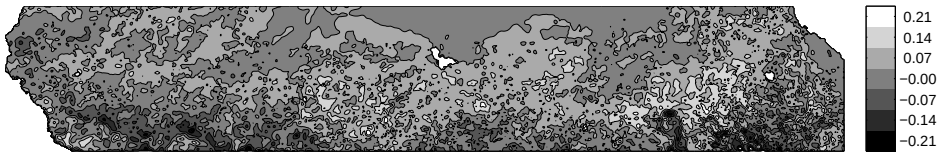

Fig. 6. Contour plot of the GMRF estimate of the $\mathbf{K}_2$ field. Here, colours corresponding to positive values correspond to areas with a positive linear increase in vegetation. Notice that a large number of these areas coincides with areas with significant positive trends in Figure 5. Also observe that this figure has a different colour scale than Figure 4.

Besides detecting purely linear trends, it is also of interest to find areas which

experience a large increase or decrease in seasonal NDVI during the first half of the time period and a lower increase or decrease during the second half. To do this, we add a third trend basis function, $f_3 = [\frac{T}{2}, \frac{T}{2} - 1, \ldots, 0, 1, \ldots, \frac{T}{2}]$, normalised to have length one. Contour plots of the resulting field, $\mathbf{K}_3$, can be seen in Figure 7. The figure contains two different contour plots, the first one shows the contours of the $\mathbf{K}_3$ field for pixels where the $\mathbf{K}_2$ field is positive. A positive value in this field means that there was a higher increase in vegetation growth during the second part of the time period, whereas a negative value corresponds to a higher increase in vegetation growth during the first part of the time period. The second plot shows the contours for the pixels where the $\mathbf{K}_2$ field is negative. Here, a positive value means that there was a larger decrease in vegetation during the first part, whereas a negative value corresponds to a larger decrease in vegetation during the second half.

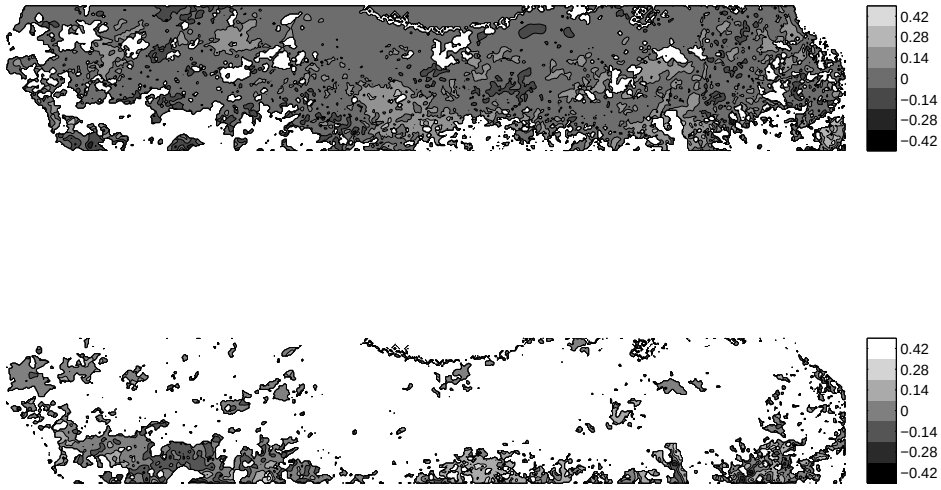

Fig. 7. Upper figure: Contour plot of the $\mathbf{K}_3$ field for the pixels where the $\mathbf{K}_2$ field is positive. A positive value in this field means that there was a higher increase in vegetation growth during the second part of the time period, whereas a negative value corresponds to a higher increase in vegetation growth during the first part of the time period. Lower figure: Contour plot of the $\mathbf{K}_3$ field for the pixels where the $\mathbf{K}_2$ field is negative. Here, a positive value means that there was a larger decrease in vegetation during the first part, whereas a negative value corresponds to a larger decrease in vegetation during the second half.

7 Extensions

In this section, we will discuss three possible extensions of the model, and the increase in complexity and potential difficulties that they would incur. A first extension is to allow for spatial dependencies in the variance of the measurement noise. Since the measurement noise includes phenomena such as cloud cover and aerosol, the model should, reasonably, allow for varying noise variances. These variances will, however, vary slowly across the image, and the measurement noise estimation could therefore be improved by adding a spatially dependent prior for the noise. There are, in principle, no problems doing this. The only difference in the model will be that explicit updating rules in the EM algorithm are hard to find, most likely necessitating numerical optimisation schemes.

Another interesting extension is to allow for varying strength in the spatial dependencies. That is, instead of using a single precision factor, κ, in the GMRF prior, we let κ vary across the field. This is probably a more accurate model for describing vegetation since the strength of the spatial dependencies most likely varies with the character of the land. Because of the problem structure, the different weights will not separate in the likelihood expression, and the resulting optimisation problem becomes highly non-linear and computationally demanding.

Finally, introducing non-Gaussian observations to handle heavy tailed residuals, or allowing generalised linear regression models, would also improve the model. However for non-Gaussian observations the expectations in (11) will most likely have to be found using simulation techniques which increases the computational burden of the algorithm.

8 Concluding Remarks

There is a need for efficient methods for estimating trends in Earth Observation data. The spatio-temporal regression model constructed in this work shows great promise for utilising the spatial dependencies in satellite-derived NDVI data. Tests on real and simulated data sets indicate that there is a substantial gain in precision compared to using independent ordinary least squares regressions for the individual pixels. By estimating the model parameters using the EM algorithm, we also obtain a substantial gain in computational cost compared to a full MCMC-based approach.

The GMRF estimates are smoother and exhibit larger contiguous regions with significant trends than a comparative analysis using OLS. The larger contigu-

ous regions and smoother estimates will most likely aid interpretation of the data and make it easier to identify underlying reasons for the detected changes in vegetation.

The model we have used assumes as common precision parameter for the entire field, which is probably not valid for the entire Sahel. Therefore, the results presented in Section 6.2 are obtained by first dividing the area into smaller subareas. This way, κ is assumed to be constant only on the smaller subareas. To avoid discontinuities in the joints between the areas, extra pixels, overlapping the neighbouring subareas, are added in the computations. Although this procedure will allow for some variation in the strength of the spatial dependencies, it does not solve the problem entirely. Hence, some improvements are anticipated to further improve the results for the Sahel region

Acknowledgement

D. Bolin and J. Lindström have been partially funded by the Swedish Foundation for Strategic Research (SSF) under grant A3 02:125, Spatial statistics and image analysis for environment and medicine.

Data used by the authors in this study include data produced through funding from the Earth Observing System Pathfinder Program of NASA's Mission to Planet Earth in cooperation with National Oceanic and Atmospheric Administration. The data were provided by the Earth Observing System Data and Information System (EOSDIS), Distributed Active Archive Center at Goddard Space Flight Center which archives, manages, and distributes this data set.

A Proof of Proposition 2

PROOF. This follows from a change of variables in the distribution $\pi(\mathbf{X}|\kappa)$:

$$\pi(\mathbf{X}|\kappa) \propto \exp(-\frac{1}{2}\mathbf{X}^\top \kappa \overline{\mathbf{Q}}\mathbf{X}) = \exp\left(-\frac{\kappa}{2}\mathbf{K}^\top \underbrace{(\mathbf{F}\otimes\mathbf{I})^\top(\mathbf{I}\otimes\mathbf{Q}_X)(\mathbf{F}\otimes\mathbf{I})}_{Q}\mathbf{K}\right).$$

Using common calculation rules (mixed product) for the Kronecker product, the expression for $\mathbf{Q}$ simplifies to

$$\mathbf{Q} = (\mathbf{F}\otimes\mathbf{I})^\top(\mathbf{I}\otimes\mathbf{Q}_X)(\mathbf{F}\otimes\mathbf{I}) = (\mathbf{F}^\top\mathbf{F})\otimes\mathbf{Q}_X. \tag{A.1}$$

Hence, $\mathbf{K}|\kappa \in \mathsf{N}(0, (\kappa\mathbf{Q})^{-1})$, and the corresponding graph can be determined from the non-zero pattern of $\mathbf{Q}$. To determine the rank of the field, we have to determine the rank of $\mathbf{Q}$. Since $\mathbf{X}_t$ is a second-order IGMRF, $\mathbf{Q}_X$ has rank $n-3$, and $\mathbf{F}^\top\mathbf{F}$ has, by construction, full rank. A basic property of the Kronecker product is that $\text{rank}(\mathbf{A}\otimes\mathbf{B}) = \text{rank}(\mathbf{A})\,\text{rank}(\mathbf{B})$, and hence, $\mathbf{Q}$ has rank $m(n-3)$. $\square$

References

Agbu, P., James, M., 1994. The NOAA/NASA Pathfinder AVHRR Land Data Set User's Manual. Goddard Distributed Active Archive Center, NASA, Goddard Space Flight Center, Greenbelt.

Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics (with discussion). Annals of the Institute of Statistical Mathematics 43, 1–59.

Bolin, D., 2007. Estimating vegetation trends in the African Sahel using Gaussian Markov random fields. Master's thesis, Lund University, Lund, LUTFMS–3090–2007.

Chappell, A., Seaquist, J., Eklundh, L., 2001. Improving the estimation of noise from NOAA AVHRR NDVI for Africa using geostatistics. International Journal of Remote Sensing 22, 1067–1080.

Cressie, N., 1991. Statistics for Spatial Data. John Wiley & Sons Ltd.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39 (1), 1–38.

Di Giacinto, V., Dryden, I., Ippoliti, L., Romagnoli, L., 2005. Linear smoothing of noisy spatial temporal series. Journal of Mathematics and Statistics 1 (4), 300–312.

Eklundh, L., Olsson, L., 2003. Vegetation index trends for the African Sahel 1982–1999. Journal of Geophysical Research 30, 1430–1433.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and Its Applications. Chapman & Hall.

Gamerman, D., Moreira, A., Rue, H., 2003. Space-varying regression models: specifications and simulation. computational statistics and data analysis 42, 513–533.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004. Bayesian Data Analysis, 2nd Edition. Chapman & Hall/CRC.

Hanan, N., Prince, S., Holben, B., 1995. Atmospheric correction of AVHRR data for biophysical remote sensing of the Sahel. Remote Sensing of Environment 51, 306–316.

Hickler, T., Eklundh, L., Seaquist, J., Smith, B., Ardö, J., Olsson, L., Sykes, M., Sjöström, M., 2005. Precipitation controls Sahel greening trend. Geophysical Research Letters 32, L21415, doi:10.1029/2005GL024370.

Holben, B. N., 1986. Characteristics of maximum-value composite images from temporal AVHRR data. International Journal of Remote Sensing 7, 1417–1434.

Hulme, M., 2001. Climatic perspectives on Sahelian desiccation: 1973–1998. Global Environmental Change 11, 19–29.

James, M., Kalluri, S., 1994. The Pathfinder AVHRR Land data set: An improved coarse resolution data set for terrestrial monitoring. International Journal of Remote Sensing 15, 3347–3363.

Jönsson, P., Eklundh, L., 2002. Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Transactions on Geosciences and Remote Sensing 40 (8), 1824–1832.

Jönsson, P., Eklundh, L., 2004. TIMESAT - a program for analyzing time-series of satellite sensor data. Computers and Geosciences 30, 833–845.

Justice, C., Hiernaux, P., 1986. Monitoring the grasslands of the Sahel using NOAA AVHRR data: Niger 1983. International Journal of Remote Sensing 7, 1475–1497.

Kidwell, K. B. (Ed.), 1998. NOAA Polar Orbiter Data User's Guide (TIROS–N, NOAA–6, NOAA–7, NOAA–8, NOAA–9, NOAA–10, NOAA–11, NOAA–12, NOAA–13 and NOAA–14) November 1998 Revision. U.S. Department of Commerce, National Oceanic and Atmospheric Administration.
URL `http://www2.ncdc.noaa.gov/docs/podug/`

Lagarias, J., Reeds, J. A., Wright, M. H., Wright, P. E., 1998. Convergence properties of the nelder-mead simplex method in low dimensions. SIAM Journal of Optimization, 112–147.

Lamb, P., 1982. Persistence of Subsaharan drought. Nature 299, 46–47.

Lindström, J., Eklundh, L., Holst, J., Holst, U., 2006. Influence of solar zenith angles on observed trends in the NOAA/NASA 8 km Pathfinder normalized difference vegetation index over the African Sahel. International Journal of Remote Sensing 27 (10), 1973–1991.

Myneni, R., Keeling, C., Tucker, C., Asrar, G., Nemani, R., 1997. Increased plant growth in the northern high latitudes from 1981 to 1991. Nature 386, 698–702.

Nicholson, S., 2000. Land surface process and Sahel climate. Reviews of Geophysics 38, 117–140.

Olsson, L., 1993. On the causes of famine – drought, desertification and market failure in the Sudan. Ambio 22, 395–403.

Olsson, L., Eklundh, L., Ardö, J., 2005. A recent greening of the Sahel–trends, patterns and potential causes. Journal of Arid Environments 63, 556–566.

Press, W., Teukolsky, W., Vetterling, W., Flannery, B., 1992. Numerical recipes in FORTRAN: The art of scientific computing. Cambridge Univ. Press.

Prince, S., 1991. Satellite remote sensing of primary production: comparison of results for Sahelian grasslands 1981-1988. International Journal of Remote Sensing 12, 1301–1312.

Prince, S., Goward, S., 1996. Evaluation of the NOAA/NASA pathfinder AVHRR land data set for global primary production modelling. International Journal of Remote Sensing 17, 217–221.

Rouse, J., Haas, R., Schell, J., Deering, D., 1973. Monitoring vegetation systems in the great plains with erts. Third ERTS Symposium, NASA, SP-351, NASA, Washington, DC, 1, 309–317.

Rue, H., 2007. GMRFLib: Fast and exact simulation of Gaussian Markov random fields on graphs (ver. 3.0-0).
URL `http://www.math.ntnu.no/ hrue/GMRFLib/`

Rue, H., Held, L., 2005. Gaussian Markov Random Fields, Theory and Applications. Chapman & Hall/CRC.

Rue, H., Martino, S., 2007. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. Journal of Statistical Planning and Inference 137 (10), 3177–3192.

Savitzky, A., Golay, M. J., 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 36, 1627–1639.

Seaquist, J., Olsson, L., Ardö, J., 2003. A remote sensing-based primary production model for grassland biomes. Ecological Modelling 169, 131–155.

Sellers, P. J., Dickinson, R. E., Randall, D. A., Betts, A. K., Hall, F. G., Berry, J. A., Collatz, G. J., Denning, A. S., Mooney, H. A., Nobre, C. A., Sato, N., Field, C. B., Henderson-Sellers, A., 1997. Modeling the exchange of energy, water, and carbon between continents and atmosphere. Science 275, 602–609.

Stowe, L., McClain, E., Carey, R., Pellegrino, P., Gutman, G., Davis, P., Long, C., Hart, S., 1991. Global disribution of cloud cover derived from NOAA/AVHRR operational satellite data. Adv. Space Res. 3, 51–54.

Tanré, D., Holben, B., Kaufman, Y., 1992. Atmospheric correction algorithm for NOAA AVHRR products: Theory and application. IEEE Trans. Geosci. Remote Sensing 30, 231–248.

Tucker, C., Vanpraet, C., Sharman, M., Van Ittersum, G., 1985. Satellite remote sensing of total herbaceous biomass production in the Sengalese Sahel: 1980–1984. Remote Sensing of Environment 17, 233–249.

Wikle, Christopher K.and Berliner, L. M., Cressie, N., 1998. Hierarchical Bayesian space-time models. Environmental and Ecological Statistics 5 (2), 117–154.